



DATA MANAGEMENT – FROM COLLECTION TO DISSEMINATION

TRIALS AND TRIUMPHS WITH DATA MANAGEMENT



Robert Stevenson, Ph.D.

Associate Professor of Biology

University of Massachusetts, Boston

robert.stevenson@umb.edu

CITIZEN SCIENCE DATA MANAGEMENT

- Data Management Context
- Data Loss
- Data Life Cycle
- Four Challenges
- Three tips



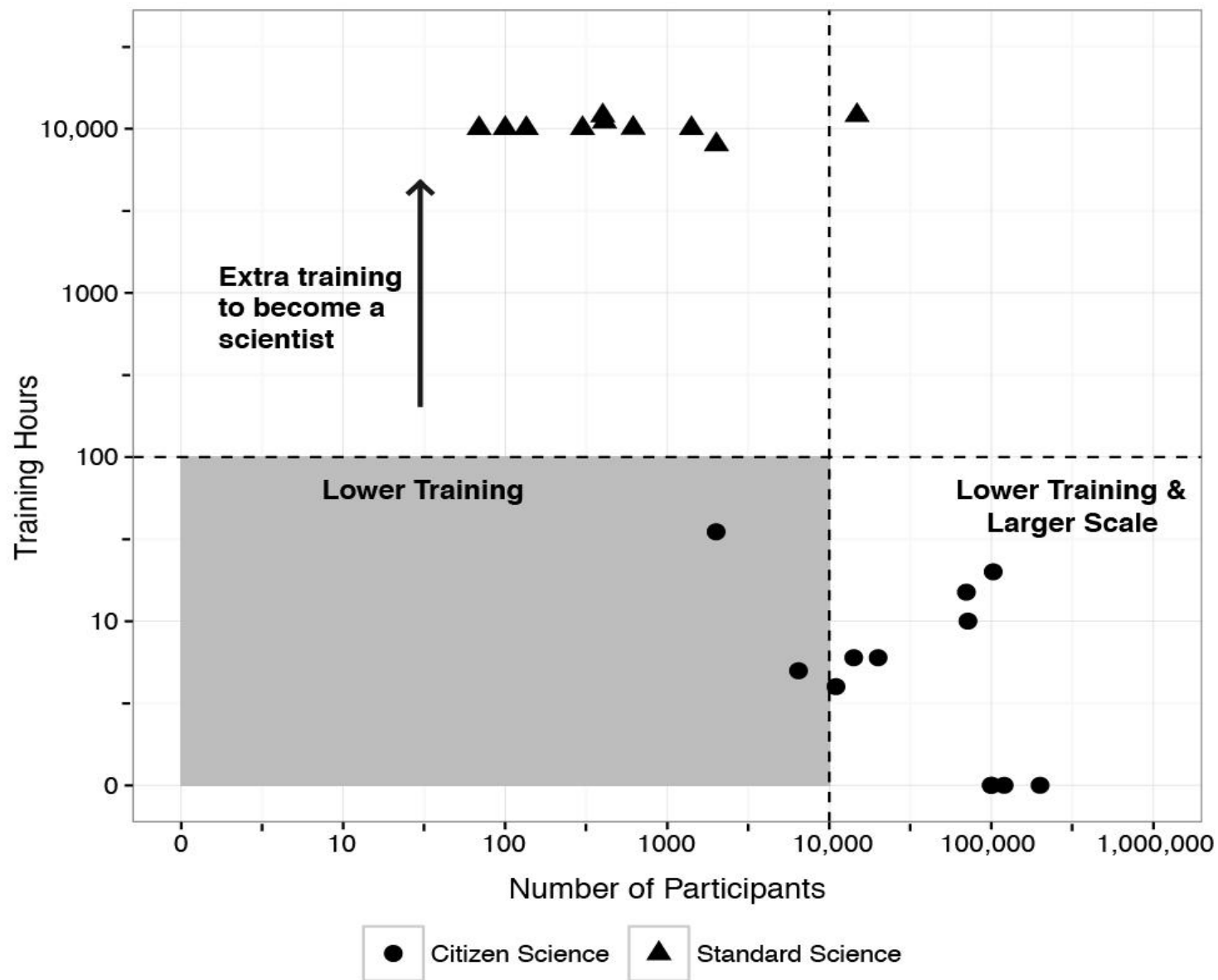
THE “SCIENCE” IN CITIZEN SCIENCE

The scientific product of citizen science projects are
DATA

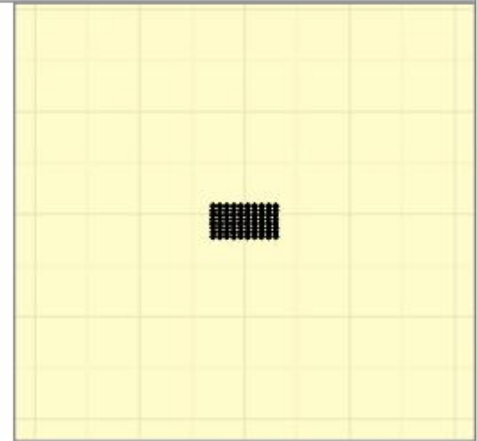
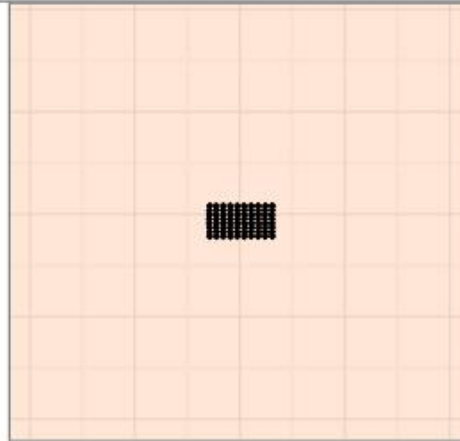
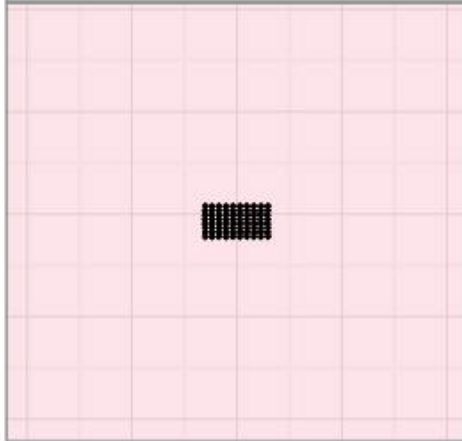
The BIG MESSAGES

Citizen Science is a new kind of Instrument
Make the data flow

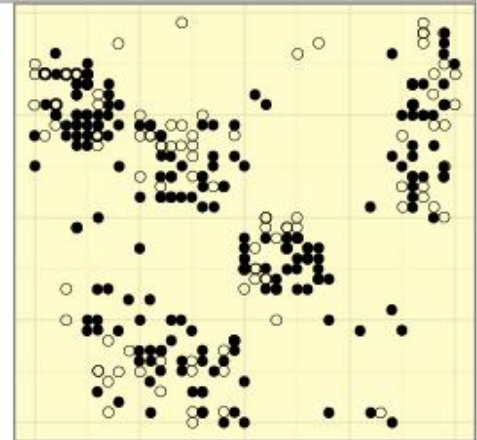
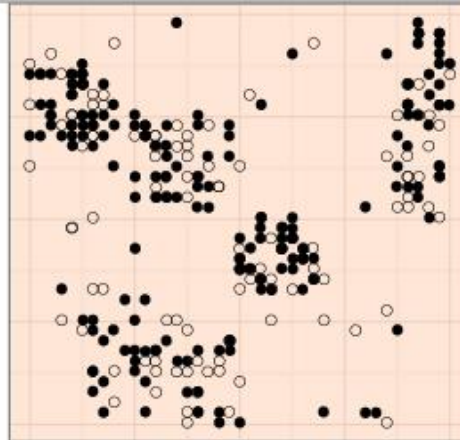
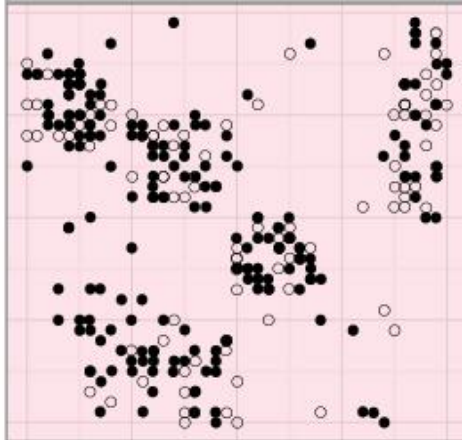




Scientific Camera



Citizen Science Camera

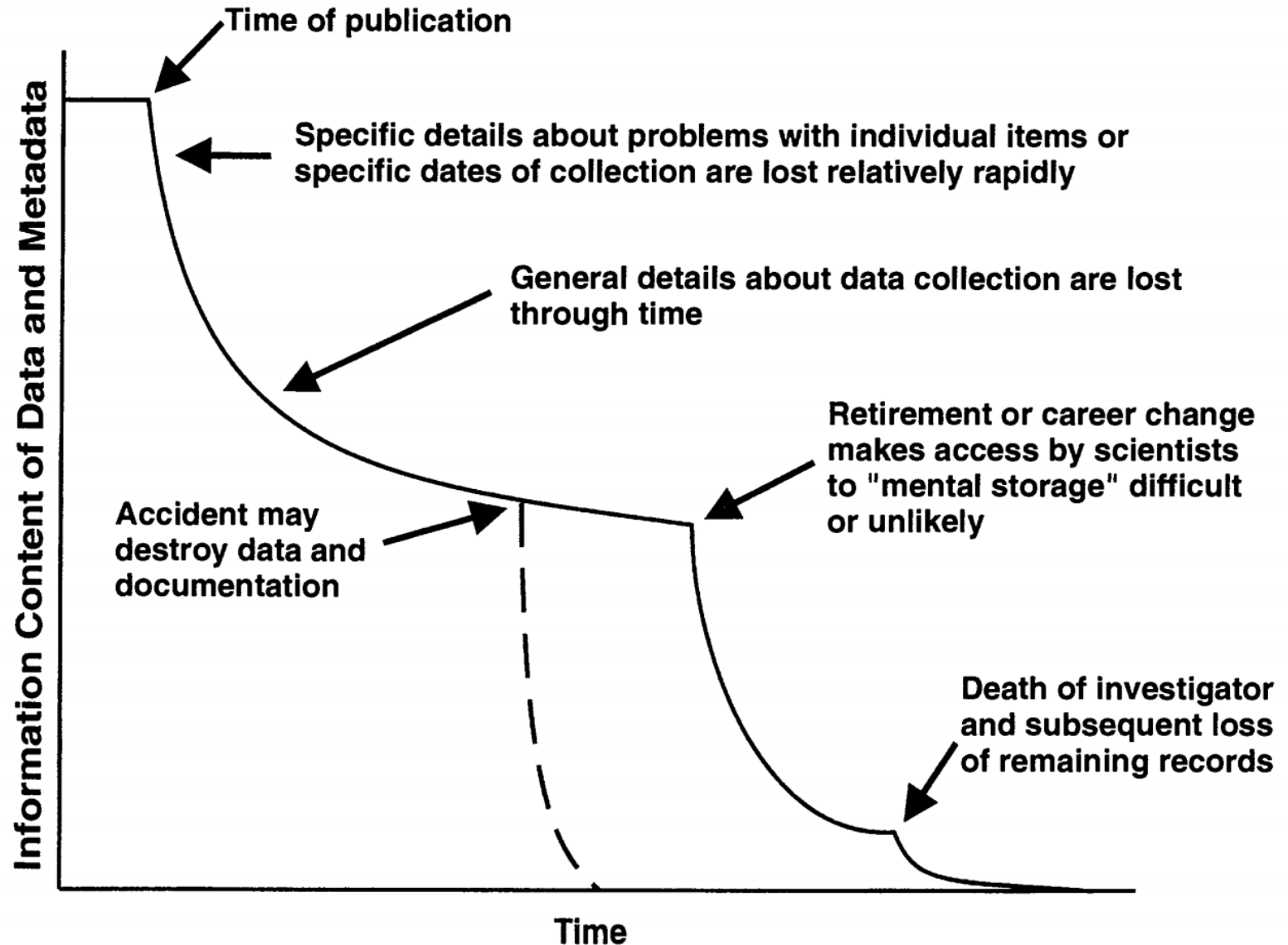


Time



Sensor Type	Example 1		Example 2		Example 3		Data Contribution	Specific data quality strategies	Papers about data quality strategies
	Project Name	Description	Project Name	Description	Project Name	Description			
Carry sensors or pilot vehicles that carry sensors, CS	TZOA, Airbeam	monitor environmental air quality	SeaKeepers International	works with NOAA and WMO and deploys Seakeeper Difters and Argo floats	Weather Underground	connects consumer weather instruments in a network	indirectly through deployment of instrument package	Calibration before and after deployment. Local and quality of instruments ranked	Bell et al (2013)
Invent or modify algorithms, IMA	Fold-It	to understand how proteins fold	MATLAB Online Programming Contest	Semi-annual event to encourage programmers to develop and share code	EteRNA	RNA folding problems	algorithms, beat the best computer algorithms	The interactive nature of the process controls data quality	None found
Sort and classify physical objects, SCPO	Passport in Time	archeology program with the USFS	Field Museum Collection Center Volunteers	count, sort and digitize artifacts and specimens	American Museum of Natural History	Volunteering in the Division of Paleontology	object categorized	Expert review, subsample comparison, blanks and standards for water analysis	Herron et al (2004), Obrecht et al (1998)
Sort and classify digital objects, SCDO	Galaxy Zoo	classify galaxies from digital images	EyeWire	map neurons in the eye of Drosophila	Old Weather	transcribe weather records from ships' logs	digital object categorized	testing with known objects, classification of real and test object by multiple citizens, statistical resolution, expert review	Lintott et al (2008), Fortson et al (2012), Willet et al (2013), Hansen et al (2011)
Collect physical objects, CPO	Florida LakeWatch	collect water samples for analysis	School of Ants	collect ants around schools that are submitted for identification	The Bighorn Basin Dinosaur Project	find and collect dinosaur fossils	sample obtained and submitted, collection process documented	replicate samples, for lab processing use splits, blanks and standards for water analysis, expert review	EPA (2002), Obrecht et al (1998), Williams (2000)
Collect digital objects, CDO	Juneau Humpback Whale Flukes	collect images of whale flukes	BatME	collect audio recordings of bats with mobile devices	PicturePost	digital images of landscape	digital object obtained and submitted, collection process documented	automate metadata collection	None found
Report observations, RO	Great Sunflower Project	Pollinator activity in gardens	CoCoRHAS	submit data about rainfall, hail events and snow fall	Feeder Watch	counts bird species that visit bird feeders	text from instrument readings, counts, classifications and/or descriptions	pseudo-replication, project specific knowledge, machine review, expert review	Kelling et al (2012; 2015b; 2015a), Yu et al (2010; 2012)

Data are lost over time





CHALLENGES

- User interface design
- Data Quality
- Metadata completeness without overwhelming participants – Use “Standards”
- Archive data



TIPS

- Match your data management solutions to scale of the project
- Become part of a larger project
- Work with a science librarian or data scientist



Making sense of the Data



Rachel Jakuba, PhD, Science Director
Buzzards Bay Coalition



Baywatchers Monitoring Program

- Program initiated in 1992 with two key goals:
 - Evaluate WQ and long term ecological health
 - Educate citizens on local WQ and environmental management issues
- ~200 stations monitored by ~130 volunteers
- DO, temp, salinity & secchi depth performed every ~5 days in the summer (late May to Sept)
- Nutrient and Chl collected 4 times a summer (2x in July, 2x in August), analyzed by MBL
- Methods approved by EPA and DEP



Our Data Management Approach

H73593 :

X

✓

f

x

Weweantic River

	B	C	E	F	G	H	K	L	M	N	O	P	Q	R	S	V	W	Y	AA	AB	AP	AR	BH	BI	BJ
	Stn_OLD	Stn_NEW	SAMP_DATE	YEAR	UniqueID	Embayment	Dataset	Source	GEN_QC	TIME	Time_QC	Dup	S_D	SamDep_m	SamDep_QC	Tot_Dep_m	TotDep_QC	DO_Meth	Sal_meth	Sal_QC	QA_Comm	BWV_Comm	SAL_FIN	DOmg_FIN	DO_Per_FIN
1																									
73588	WW6	6/25/2015	2015	WW6-X-2015-6-25-645-S-0.15-0	Weweantic River	X	BBC			6:45 AM		0	S	0.15		0.30	DO2	S2				filamentous gre	10.60	7.06	87.29%
73589	WW6	7/1/2015	2015	WW6-X-2015-7-1-630-S-0.15-0	Weweantic River	X	BBC			6:30 AM		0	S	0.15		0.75	DO2	S2				high water-tan	13.12	6.02	75.70%
73590	WW6	7/1/2015	2015	WW6-X-2015-7-1-632-D-0.5-0	Weweantic River	X	BBC			6:32 AM		0	D	0.5		0.75	DO2	S2				high water-tan	15.16	5.95	75.71%
73591	WW6	7/6/2015	2015	WW6-X-2015-7-6-639-S-0.1-0	Weweantic River	X	BBC			6:39 AM		0	S	0.1		0.15	DO2	S2					6.43	6.09	71.35%
73592	WW6	7/13/2015	2015	WW6-N-2015-7-13-116-S-0.15-1	Weweantic River	N	BBC			11:06 AM		1	S	0.15		0.30	DO2	S2				wade in	7.13	7.92	104.06%
73593	WW6	7/13/2015	2015	WW6-N-2015-7-13-117-S-0.15-1	Weweantic River	N	BBC			11:07 AM		1	S	0.15		0.30	DO2	S2				wade in	7.13	7.92	104.06%
73594	WW6	7/13/2015	2015	WW6-N-2015-7-13-00-S-0.15-2	Weweantic River	N	BBC					2	S	0.15		0.30	DO2	S2				wade in	7.13	7.92	104.06%
73595	WW6	7/16/2015	2015	WW6-X-2015-7-16-641-S-0.15-0	Weweantic River	X	BBC			6:41 AM		0	S	0.15		0.50	DO2	S2					14.09	5.08	63.35%

<

>

Readme

all

META

Acknowledge

Stations

Methods

ChangeLog

EquivStns

⊕

:

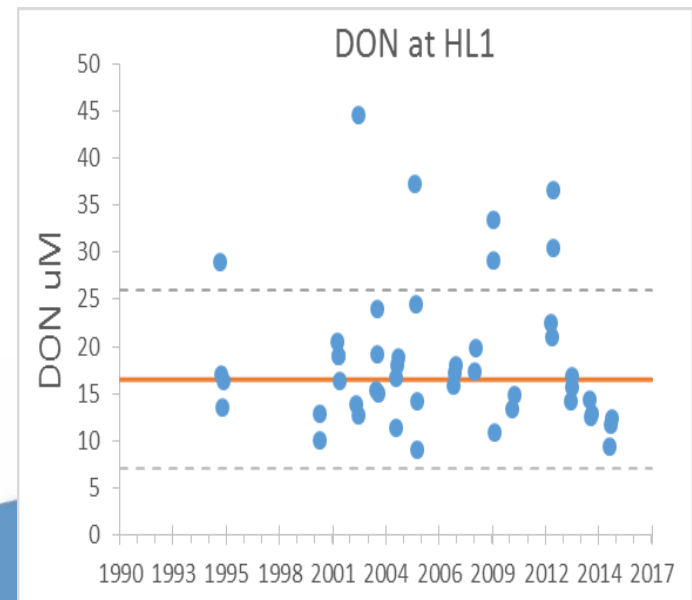
<

>

- Microsoft Excel structured reference table –
 - Standard program that staff have access to and are familiar with
 - Easily sorted by any column parameter
 - Can be exported as a CSV file for use by other programs
- Metadata and QA data incorporated in one place

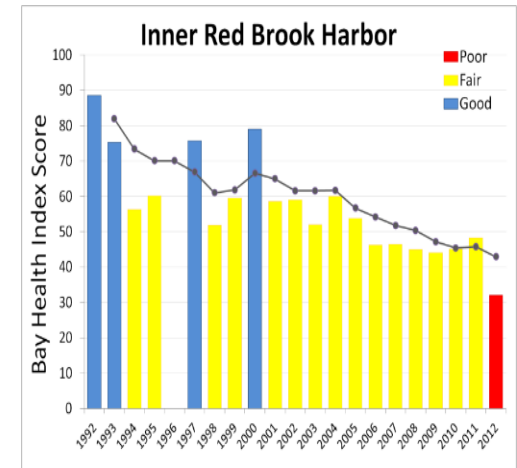
How/when do we QA our data?

- During data entry
- After all volunteer data is entered
- When new Lab data received
- After new year's data is combined with previous years
- During Bay Health Index calculation



How do we use our data?

- Track trends in nutrient pollution impacts
- Educate the community on Bay health
- Provide regulators with critical info on WQ
- Give students a real-world application of science



● Good to Excellent (65-100)
● Fair (35-65)
● Poor/Eutrophic Conditions (<35)



Calculating the Bay Health Index

- Key monitoring parameters are combined to produce a Bay Health Index score on a scale of 0 to 100

- Good to Excellent (65-100)
- Fair (35-65)
- Poor/Eutrophic Conditions (<35)

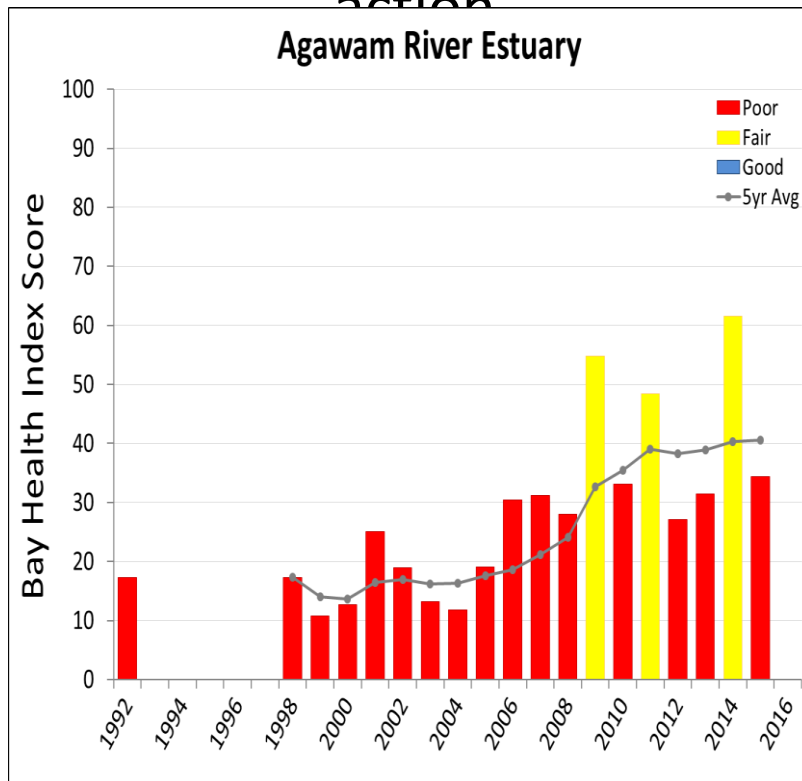
Calculating the Bay Health Index

- Key monitoring parameters are combined to produce a Bay Health Index score on a scale of 0 to 100

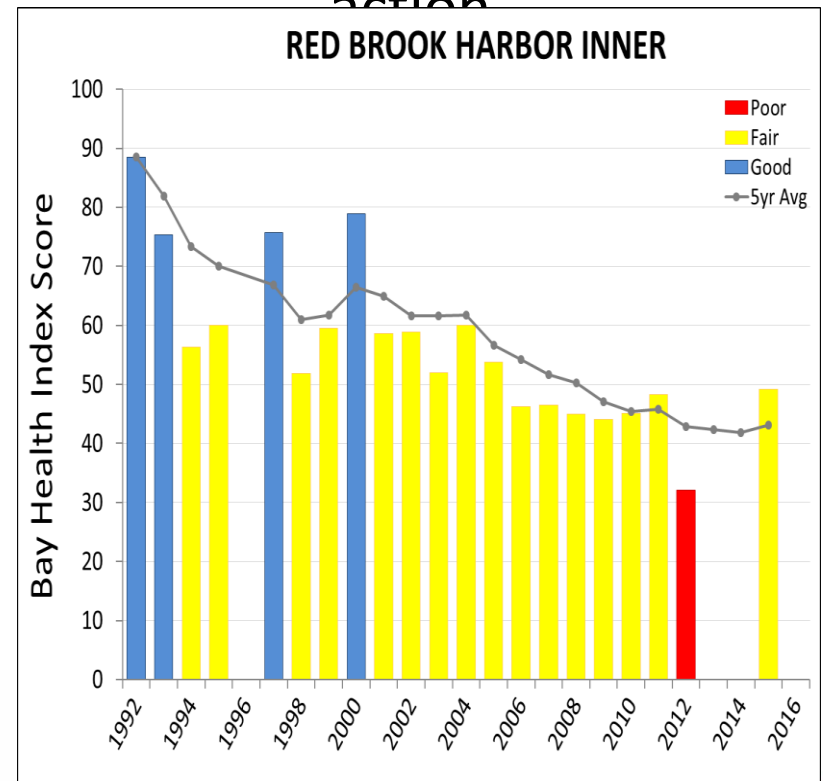
Parameter	0 points	100 points
DO (lowest 20% of measmts)	40% sat	85% sat
Secchi disk depth	0.6 m	3 m
Chla + pheo	10 $\mu\text{g L}^{-1}$	3 $\mu\text{g L}^{-1}$
DIN	10 μM (0.14 ppm)	1 μM (0.014 ppm)
TON	43 μM (0.6 ppm)	20 μM (0.28 ppm)

Tracking trends in bay health

Water quality
improvements after
action



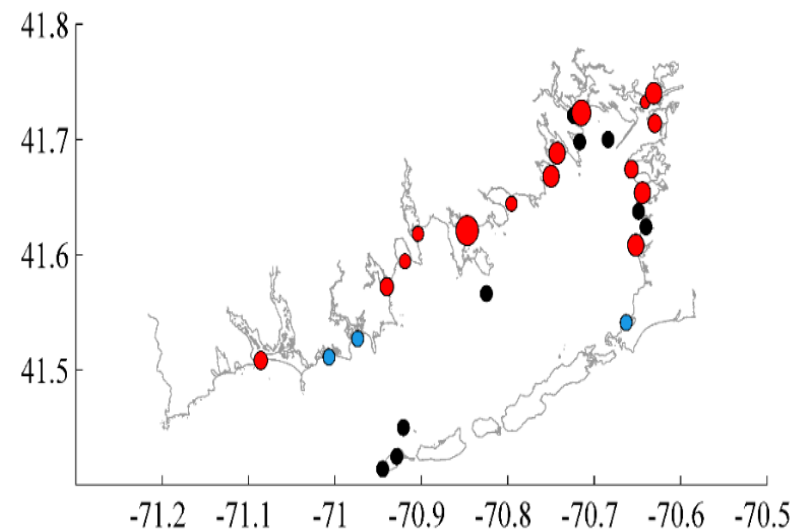
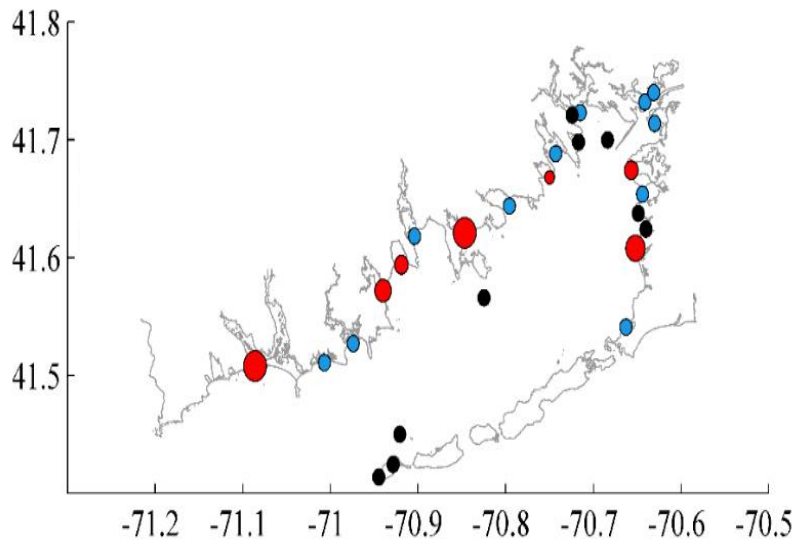
Water quality
declines with no
action



Diving Deeper



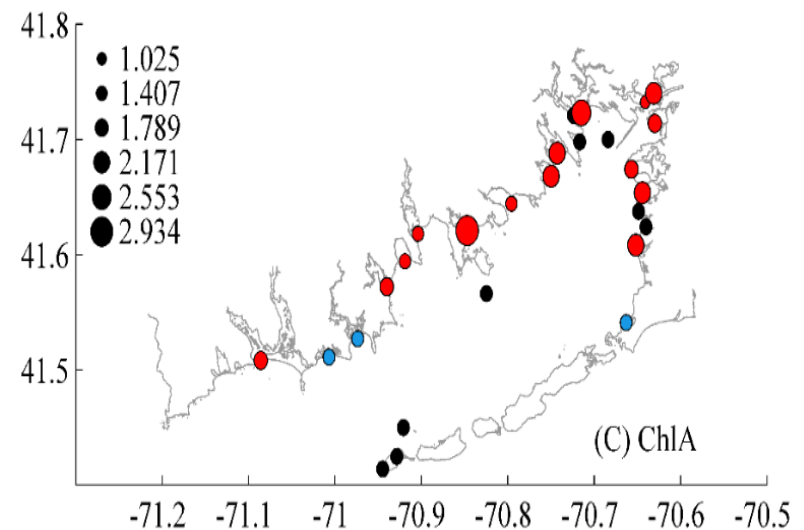
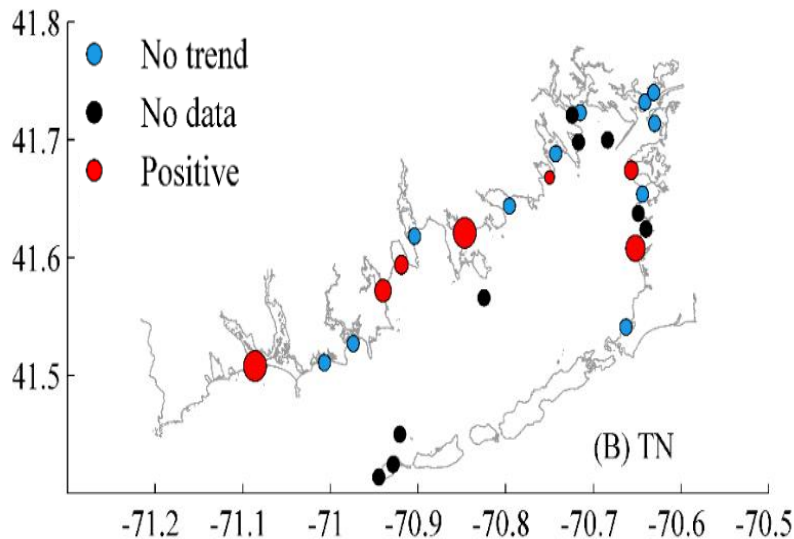
- With MacArthur Foundation grant, we are working with WHOI scientists to explore trends in the data and determine where and how climate change impacts are influencing these trends



Diving Deeper



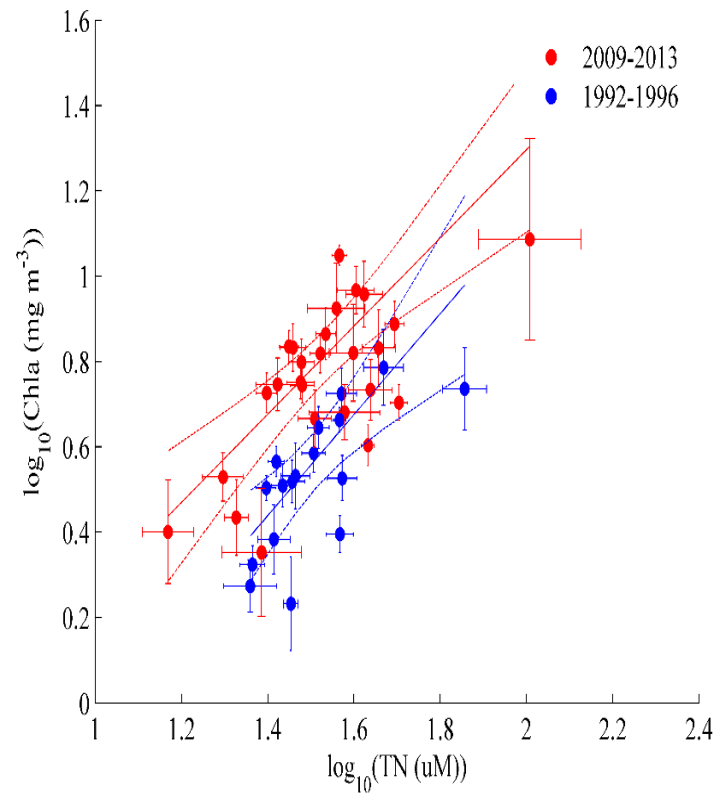
- With MacArthur Foundation grant, we are working with WHOI scientists to explore trends in the data and determine where and how climate change impacts are influencing these trends



- Red dots are where TN or ChlA significantly increased over 23 yrs
- Dot size represents magnitude of the slope

Rheuban et al.,
2016

Shift in TN and Chl relationship



Keys to Success

- Measure consistent parameters in a rigorous way
- Consider who your audience is and what information you are trying to relay
- Be organized and record as much information as possible in your database
- Look for partners with skills in data analysis (if you build it, they will come)



MAKING DATA USEFUL: SHARING RESULTS WITH STAKEHOLDERS AND POLICY MAKERS

**Andy Hrycyna
Mystic River Watershed Association**

**MassBays Citizen Monitoring Coordinators' Summit
September 29, 2016**

Interesting data

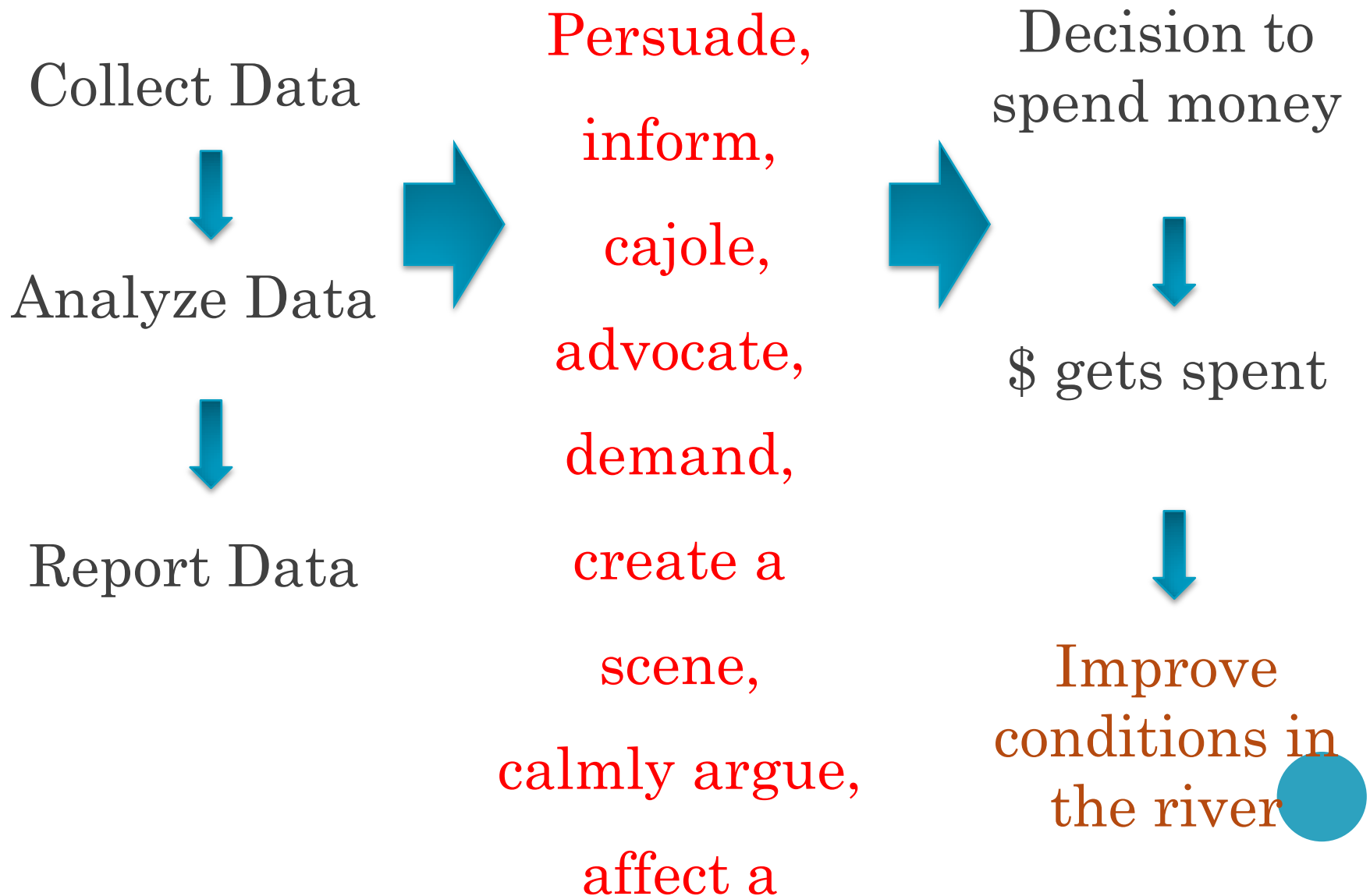
vs.

Useful data

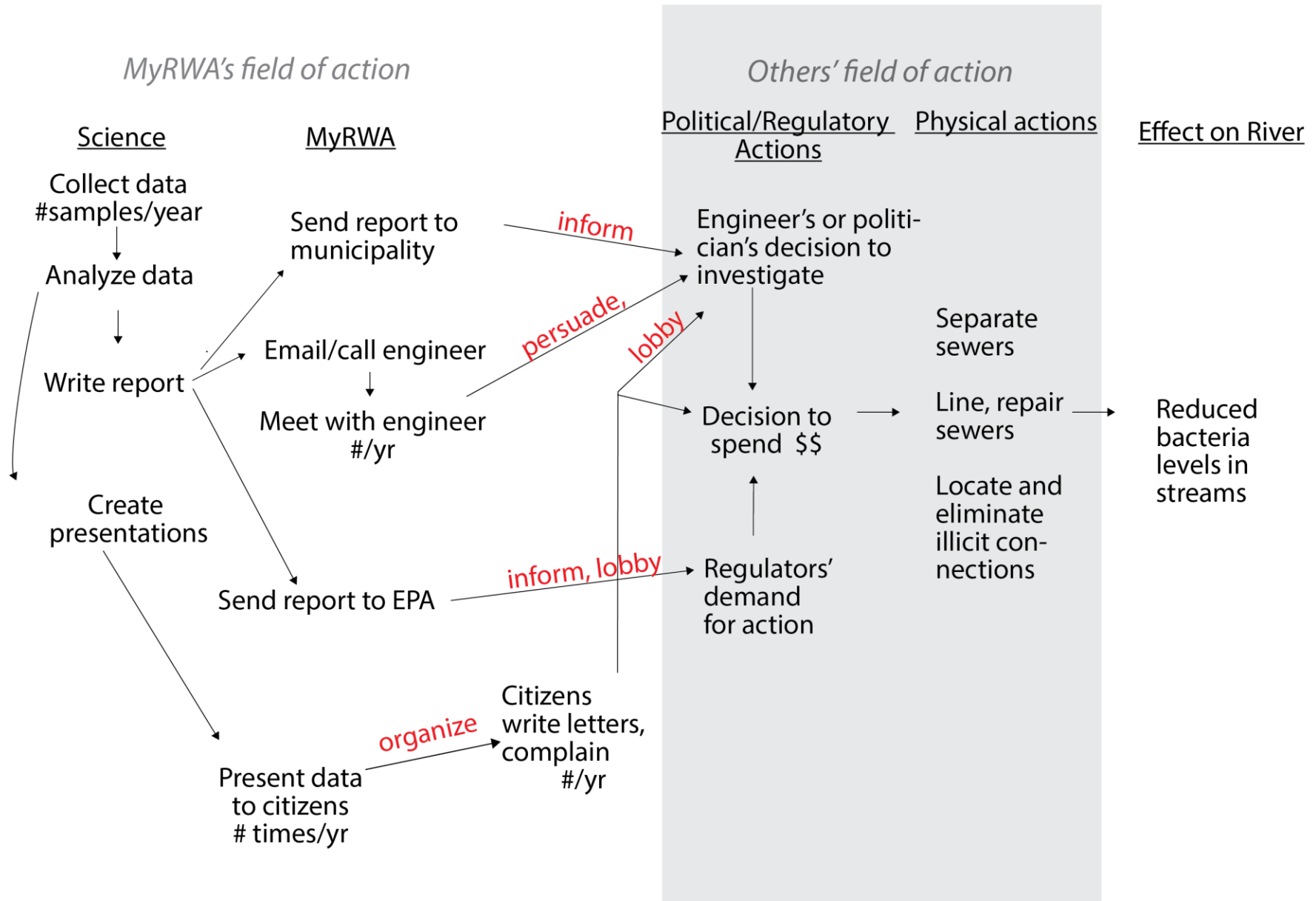
*Useful water quality data is
that which helps bring about
positive environmental change*



From Data to Change



How MyRWA can reduce bacteria levels in streams



Some ways we share/publish our data

- Website
- EPA annual water quality report card
- Hotspot reports
- Reports to collaborators





HOME WATERSHED INFO PROJECTS & PROGRAMS EVENTS PUBLICATIONS MAKE A DIFFERENCE ABOUT US

I LOVE THIS STUFF: MORE WATER QUALITY INFORMATION!

Use this page to explore more detailed data gathered by the Mystic Monitoring Network.

2016 2015 2015 Lines 2014 2014 Lines 2013 2013 Lines

2015 Raw Data from Mystic Baseline sites

Select a characteristic from the drop-down menu to view the results for each month.

Learn more about these characteristics and sampling dates at the [Monitor Resources](#) page.

Choose 2013 or 2014 tab for more data

Values, low to high

10 24,200

Characteristic Name

Escherichia coli

2015 Escherichia coli (MPN/100ml)

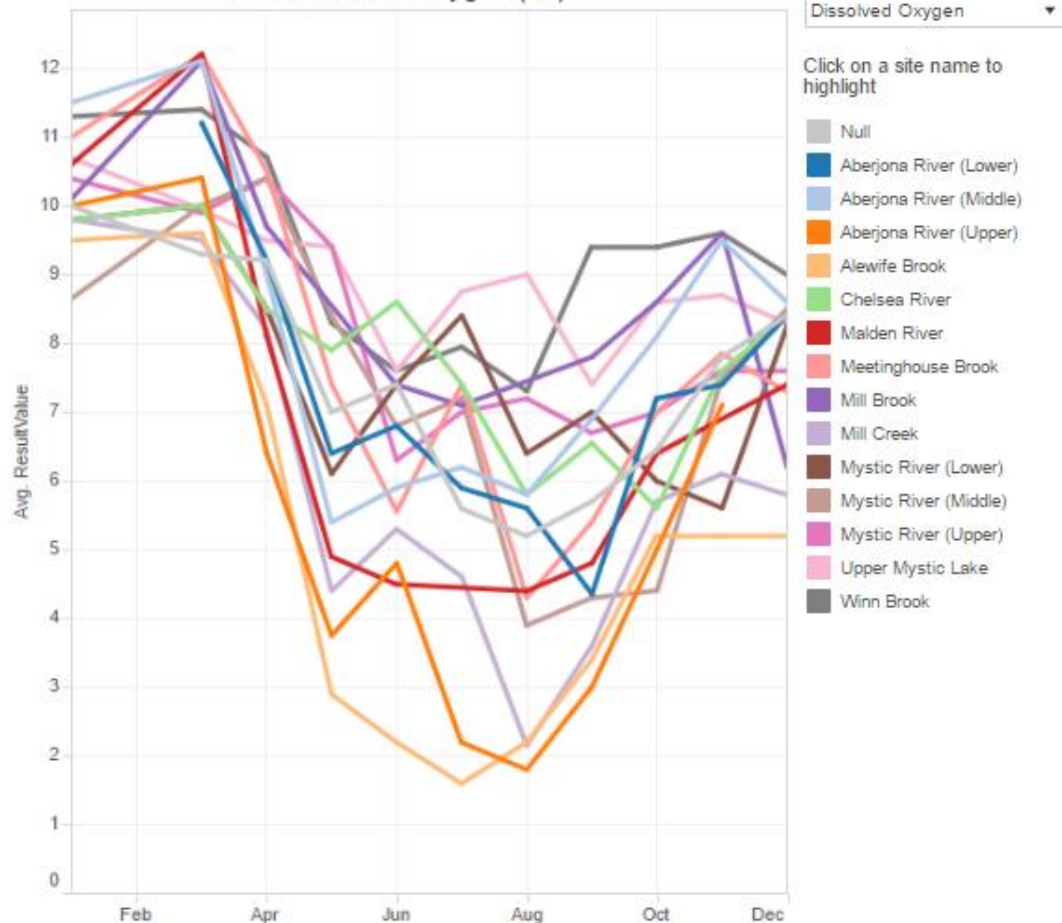
	January	March	April	May	June	July	August	September	October	November	December
Aberjona River (Lower)	213	249	92	8,630	842	404	547	170	52	10	935
Aberjona River (Middle)	135	259	35	17,030	639	1,150	24,200	313	458	52	832
Aberjona River (Upper)	63	426	34	212	327	3,260	9,800	495	1,630	121	
Alewite Brook	134	504	134	278	185	209	428	450	318	233	830
Malden River	1,450	1,370	110	5,170	973	63	36	135	118	128	4,110
Meetinghouse Brook	15	120	10	240	63	72	148	613	712	136	703
Mill Brook	173	231	102		1,330	1,200	700	609	530	199	435
Mystic River (Upper)	10	10	20	211	350	187	74	10	31	51	496
Upper Mystic Lake	10	52	10	63	63	10	10	10	10	10	37
Winn Brook	368	179	413	1,630	1,700	2,955	1,550	933	6,490	496	2,275

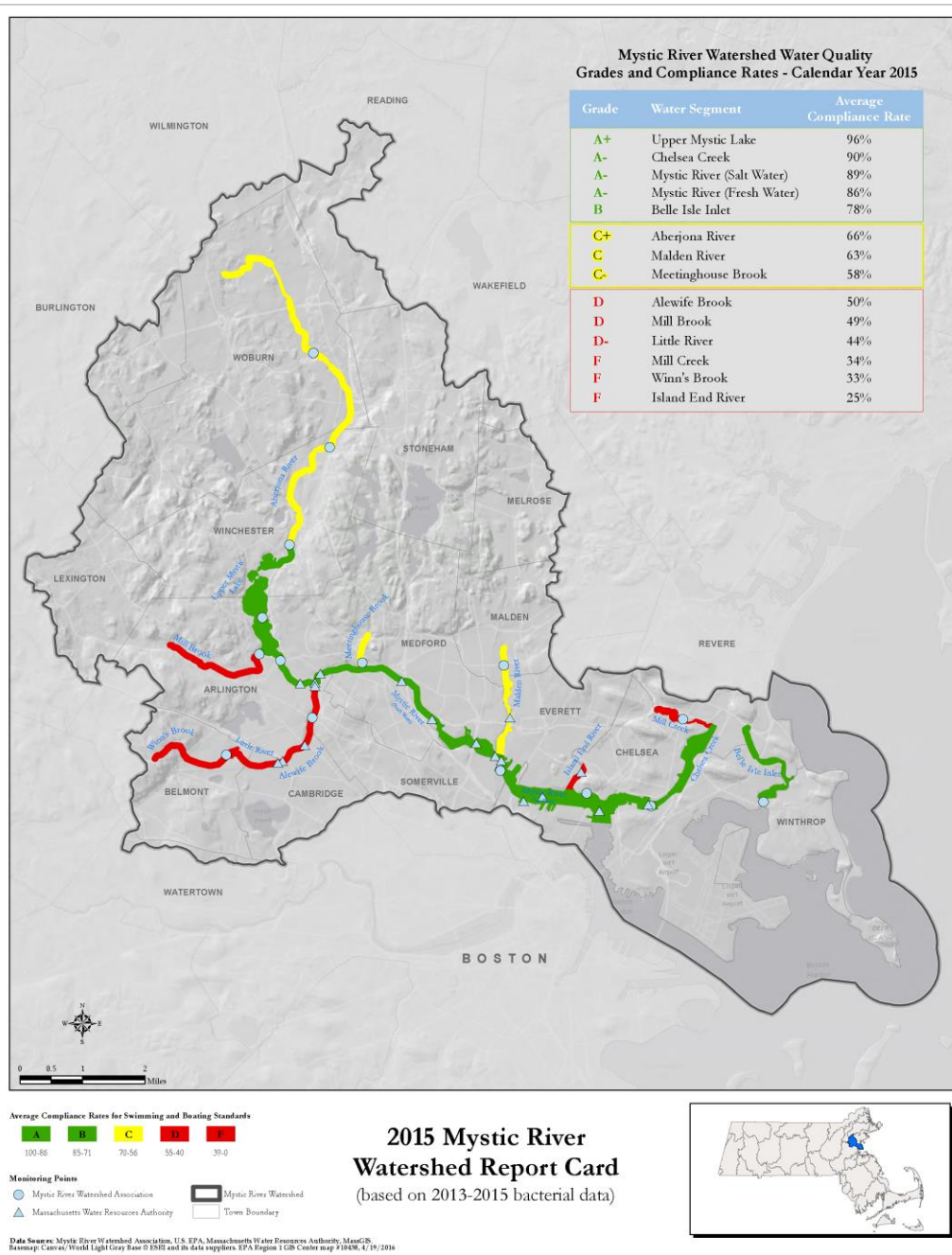
I LOVE THIS STUFF: MORE WATER QUALITY INFORMATION!

Use this page to explore more detailed data gathered by the Mystic Monitoring Network.

2016 2015 2015 Lines 2014 2014 Lines 2013 2013 Lines

2015 Dissolved Oxygen (All)





EPA water quality report card

- MyRWA does analysis
- MyRWA + MWRA data sets
- Relatively simple calculations of %'s
- Even simpler, intuitive grading system
- Power of maps



EDITORIAL

Another year, another 'F' for Belmont water quality



14



ALAN WITZBICKI/GLOBE STAFF

Winn's Brook in Belmont once again got an 'F' on its environmental report card.

JUNE 03, 2016

REALLY, BELMONT? Another year, another [failing grade](#) for a key point in Belmont's drainage

Top 10 Trending Articles

Most Viewed

Most Commented

Most Shared

Federal probe finds harassment, discrimination at Boston Latin School, orders reforms

For Maine restaurateur, a gun debate turns nasty

Christopher L. Gasper: The rest of the NFL doesn't get it: Why the Patriots succeed

Boston Latin School to lose top standing in latest test scores

This is classic Rex Ryan

Sustainable infrastructure after the Automobile Age

Man searched for his mother after boat sank in Atlantic, Coast Guard says

Well, well, well ... look what the drought's doing now

At debate, Trump brings the crowd. Can Clinton win them over?

Can the Red Sox carry momentum into the playoffs?

Hotspot reports



Bacterial Assessment: November 17, 2014

Water Bodies: Mystic River (Fresh), Winter Hill Brook

Municipalities: Medford, Somerville

Hotspot Sampling Date: November 17, 2014

Precipitation Record: Wet (0.99" 48 hr prior to sampling)

Report Date: January 30, 2015

Executive Summary

The Mystic River Watershed Association conducted wet-weather monitoring in the municipalities of Medford and Somerville on November 17, 2014. The main goal of the monitoring event was to follow up on high levels of bacteria observed in an October 23, 2014 survey by revisiting many of the same sites in another storm event. Water quality samples were taken in accordance with the MyRWA MA-DEP approved QAPP along with accurate notes.

Testing revealed that 11 of 12 tested sites had *E. coli* values that exceeded MassDEP Water Quality Standards for Class B water bodies for swimming. 8 of 12 sites exceeded the standard for boating as well. The highest *E. coli* values were found at Torbert MacDonald Park (MEDOF04800, 16,580 *E. coli* MPN / 100 ml) and at Winter Hill Brook (MEDOFWB010 and WNB04, 198,630 and 111,990 *E. coli* MPN / 100 ml, respectively).

MassDEP Water Quality Standards for Class B water bodies are 1260 *E. coli* / 100 ml for boating and 235 *E. coli* / 100 ml for swimming.

Figure 1: Bacteria Levels Measured on November 17, 2014

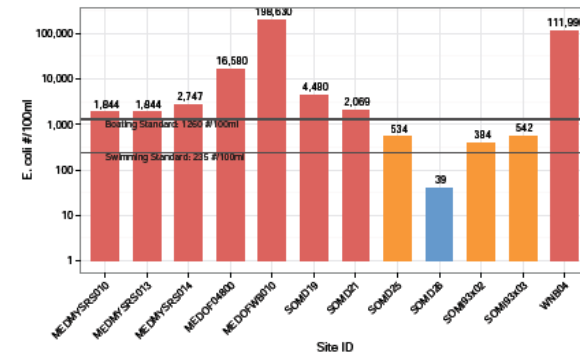


Table 1: Data collected at outfalls during Hotspot sampling event on November 17, 2014

Site ID	E. coli (#/100ml)	Surfact. (ppt)	NH3 (mg/l)	Water Body	Comment
MEDMYSRS010	1,844	0.10	0.25	Mystic River (Fresh)	pipe off wall
MEDMYSRS013	1,844	0.10	0.50	Mystic River (Fresh)	
MEDMYSRS014	2,747	0.25	1.00	Mystic River (Fresh)	
MEDOF04800	16,580	0.05	0.50	Mystic River (Fresh)	
MEDOFWB010	198,630	0.75	1.00	Winter Hill Brook	

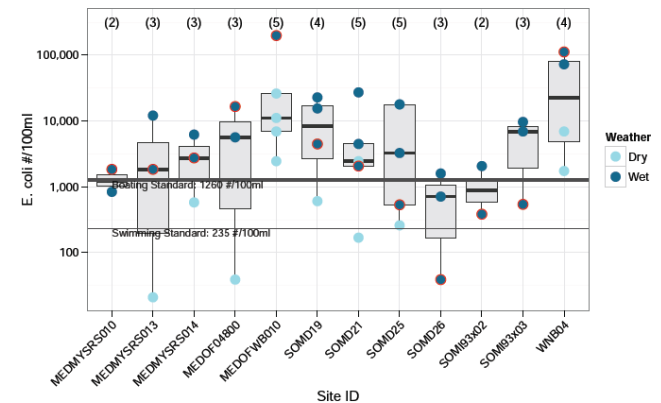


Map 2. Sampling locations in Somerville, near Blessing of the Bay Boathouse and I-93.

- Exceeds boating standard, >1260 MPN/100 ml
- Exceeds swimming standard, >235 MPN/100ml
- Meets swimming and boating standards, <235 MPN/100ml

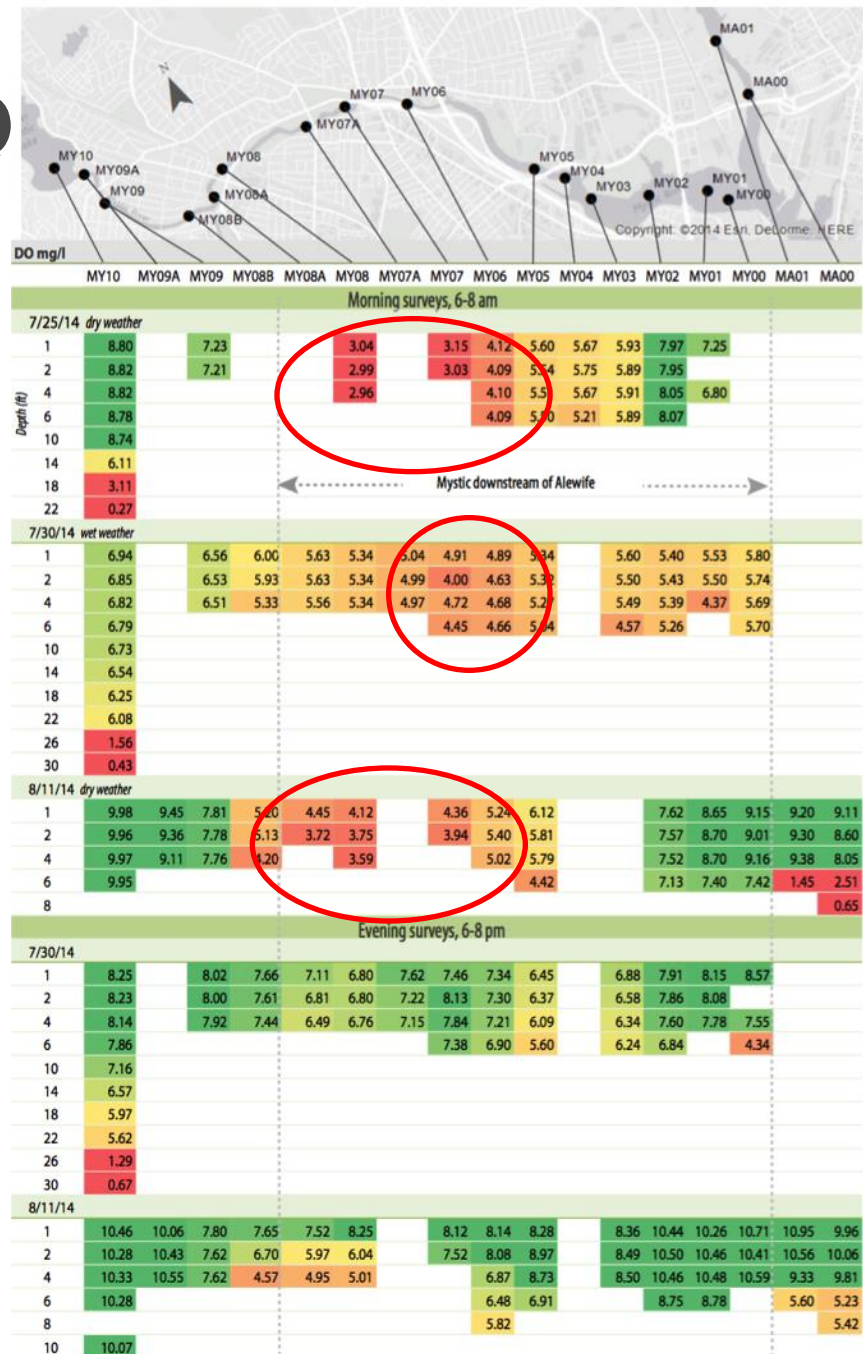
Appendix 2: Historical Data

Figure 2: Distribution of historical bacteria data at outfalls sampled during Hotspot event on November 17, 2014



2014 Mystic River surveys

- Low DO (< 5 mg/L) observed downstream of Alewife confluence



EVIDENCE OF IMPAIRMENT: CHLOROPHYLL-A

